Multi-perspective label based deep learning framework for cerebral vasculature segmentation in whole-brain fluorescence images

Yuxin Li,[1] [iD] Tong Ren,[1] Junhuai Li,[1] Xiangning Li,[2,3] and Anan Li[2,3,*]

[1]*Shaanxi Key Laboratory of Network Computing and Security Technology, School of Computer Science and Engineering, Xi'an University of Technology, Xi'an, 710048, China*
[2]*Britton Chance Center for Biomedical Photonics, Wuhan National Laboratory for Optoelectronics, MoE Key Laboratory for Biomedical Photonics, School of Engineering Sciences, Huazhong University of Science and Technology, Wuhan, 430074, China*
[3]*HUST-Suzhou Institute for Brainsmatics, Suzhou, 215123, China*
[*]aali@hust.edu.cn

Abstract: The popularity of fluorescent labelling and mesoscopic optical imaging techniques enable the acquisition of whole mammalian brain vasculature images at capillary resolution. Segmentation of the cerebrovascular network is essential for analyzing the cerebrovascular structure and revealing the pathogenesis of brain diseases. Existing deep learning methods use a single type of annotated labels with the same pixel weight to train the neural network and segment vessels. Due to the variation in the shape, density and brightness of vessels in whole-brain fluorescence images, it is difficult for a neural network trained with a single type of label to segment all vessels accurately. To address this problem, we proposed a deep learning cerebral vasculature segmentation framework based on multi-perspective labels. First, the pixels in the central region of thick vessels and the skeleton region of vessels were extracted separately using morphological operations based on the binary annotated labels to generate two different labels. Then, we designed a three-stage 3D convolutional neural network containing three sub-networks, namely thick-vessel enhancement network, vessel skeleton enhancement network and multi-channel fusion segmentation network. The first two sub-networks were trained by the two labels generated in the previous step, respectively, and pre-segmented the vessels. The third sub-network was responsible for fusing the pre-segmented results to precisely segment the vessels. We validated our method on two mouse cerebral vascular datasets generated by different fluorescence imaging modalities. The results showed that our method outperforms the state-of-the-art methods, and the proposed method can be applied to segment the vasculature on large-scale volumes.

1. Introduction

The brain relies heavily on blood vessels for a constant supply of oxygen and nutrients [1,2]. Many neurological diseases have a vascular component, such as Alzheimer's disease (AD), Parkinson's disease (PD), and stroke [3–5]. Accurate mapping of the cerebrovascular network structure is essential for understanding brain function and the pathogenesis of brain diseases. With the development of fluorescent labelling technology and imaging techniques and optical imaging, mesoscopic whole-brain optical imaging, such as fluorescence micro-optical sectioning tomography (fMOST), light-sheet microscopy (LSM) and serial two-photon microscopy (STP), could obtain an entire mouse brain vasculature at capillary-resolution [6–8]. Segmentation of

cerebral vessels from the generated image data is a fundamental step in analyzing the vessel structures and revealing the pathogenesis of brain diseases.

Unlike other image segmentation tasks, accurate segmentation of whole-brain fluorescent vascular images is challenging (see Fig. 1). First, high-resolution whole-brain imaging can obtain a fine vascular network including arteries, veins, and capillaries at the same time. The features of different scale vessels vary significantly in shape, brightness, and density. Since the fluorescent markers are labelling the epithelial cells of the blood vessels, it leads to the cavities in the centre of the thick vessels when imaging. In contrast, the fine vessels are almost no cavities. The signal-to-noise ratio in the images varies greatly due to the fluorescent signal intensities. In some micro-vessel regions, the low intensity of the fluorescence signal makes the image contrast between the vessels and the background is low. Second, the imbalance problem is serious in 3D vascular images. More than 90% of vessels in whole-brain fluorescent images are capillaries [6]. The number of features in the thick and very thin vessel regions is very sparse, resulting in the problem of being detected by segmentation methods. Last, the cerebral vasculature is a network-like structure whose topological features are biologically meaningful, so the topological structures of vascular networks should be preserved as much as possible during segmentation [9].
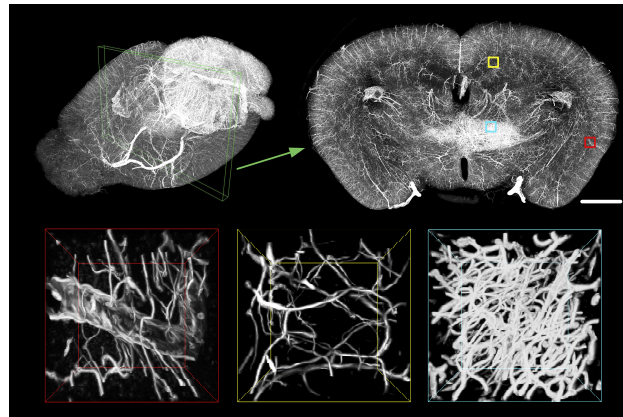
Fig. 1. Whole mouse brain vasculature image acquired by fluorescence Micro-Optical Sectioning Tomography. Three 3D volumes extracted from the entire brain data illustrate the complex features of the cerebrovascular. Scale bar, 1000 μm (the coronal sections).

Vessel segmentation is one of the most common tasks in biomedical image analysis. Many methods have been proposed for vessel segmentation in recent decades [10–12]. Existing vessel segmentation algorithms can be roughly classified into traditional vascular segmentation algorithms and deep learning-based segmentation algorithms according to whether the manually labelled ground truth is used or not.

Traditional segmentation algorithms mainly include threshold-based methods, filter-based methods, model-based methods, active contour methods, etc. For example, Frangi et al. [13] proposed a Hessian matrix-based edge detection filter to detect tubular-like structures. Zhao et al. [14] proposed a weighted symmetry filter for the segmentation of 3D vessels from different imaging modalities. Zeng et al. [15] used an anisotropic diffusion filter combined with graph cuts to segment the liver vessels. Both Shang et al. [16] and Cheng et al. [17] used the active contour models to segment the vascular tree structures. A multi-model-based approach for segmenting blood vessels in 3D fluorescence cryomicrotome images was introduced by Goyal et al. [18]. Ji et al. [8] utilized a series of filters to segment the whole-mouse-brain vessels acquired by STP. These traditional algorithms require a hand-crafted feature extractor, relying on rich prior knowledge. In addition, the designed feature extractors can only be used for specific imaging

modalities. It is challenging to design a model covering all vessels' features for whole-brain fluorescence images with complex features.

In contrast to traditional methods that utilize hand-crafted features, deep learning-based methods can automatically extract complex features from images. Benefiting from the powerful feature extraction capabilities, deep learning-based methods have been successfully used in biomedical image processing. In recent years, several excellent deep learning-based vessel segmentation methods have been proposed [12]. Convolutional neural networks (CNNs) are the most commonly used deep learning-based vessel segmentation methods. For example, Tahir et al. [19], Haft-Javaherian et al. [20] and Damseh et al. [21] used CNNs for segmentation of brain vasculature in two- or multi-photon images. Kirst et al. [7] and Todorov et al. [22] used CNN to segment mouse brain vessels imaged by light-sheet microscopy. Tetteh et al. [23] introduced DeepVesselNet, which uses 2.5D CNN instead of 3D CNN to segment vessels in 3D angiographic volumes. These deep learning methods follow the idea of semantic segmentation in computer vision. The vessels are treated as a common foreground in images, and the neural networks are used to predict the probability that each pixel belongs to the foregrounds. Recently, some specifically optimized neural networks for the segmentation of vascular targets have been proposed. Yan et al. [24], Li et al. [25] and Wang et al. [26] used the attention mechanism to design networks to focus on details or difficultly segmented vessel regions in images. Wang et al. [27] and Yan et al. [28] used multi-task approaches to separately segment different types of vessels (arteries or veins, thick or thin vessels) in images. In addition, Yan et al. [29] and Shit et al. [9] proposed skeletal similarity metrics or loss functions for the segmentation of tubular structures from a perspective of preserving the topology structure of vessels. These optimized methods require special loss functions or additional networks to complete the segmentation tasks, increasing the difficulty of network design.

Although deep learning-based methods have achieved state-of-the-art results, no single segmentation method can be applied to different imaging modalities [10], especially whole-brain fluorescence imaging. The existing deep learning methods train and segment vessels using a single type of annotated label with a per-pixel loss that the weight of each pixel is the same. However, an extremely unbalanced ratio exists among vessels with different diameters in whole-brain fluorescence images. The proportion of thick vessels and very thin vessels are small. This imbalance causes the pixel losses of thick and very thin vessels to be overwhelmed by the loss of other vessel pixels when the network is trained. The network has difficulty learning the features (cavities, low fluorescent signal) at the thick and very thin vessels, resulting in inaccurate segmentation.

To address this problem, we propose a three-stage deep learning cerebral vasculature segmentation framework based on multi-perspective labels. The segmentation task is divided into three stages. Each stage is implemented by separate models. To train these models, two labels highlighting the central region of the thick vessels and the skeleton region of vessels were generated based on the binary annotated labels by morphological operations. The first two models were trained by these two generated labels, respectively, and pre-segmented the thick vessels and the skeleton of vessels. As the labels highlight the central region and the skeleton region of vessels, the networks could pay more attention to learning the features of these regions. After pre-segmentation, the third model fuses the pre-segmented results to precisely segment the vessels. Experimental results on two different whole-brain fluorescence datasets demonstrate that our proposed methods achieved the state-of-the-art level for segmenting cerebral vasculature.

2. Materials and method

2.1. Whole-brain vasculature fluorescence imaging and dataset preparation

In this work, we used two types of fluorescent vascular data. The fluorescent dye-labelled and the transgenic-labelled whole mouse brain vessels, respectively. All animal experiments were

approved by the Institutional Animal Ethics Committee of Huazhong University of Science and Technology (the IACUC protocol number is 2017-S166 and it is valid from Sep 1, 2017 to Aug 31, 2022).

Dataset 1 is the vasculature of a C57BL/6 adult mouse imaged by fluorescence Micro-Optical Sectioning Tomography (fMOST) (BioMapping 5000; OEBio Inc.) [30]. First, the mouse was intravenously injected with DyLight594-Lycopersicon esculentum lectin, which marked vessels throughout the mouse's body. Then the mouse brain was removed and imaged continuously using the fMOST at a resolution of $0.35\mu m \times 0.35\mu m \times 2\mu m$ voxels. The size of the uncompressed image dataset exceeded 3.8 terabytes (8-bit pixel depth) with more than 5,000 coronal sections.

Dataset 2 is the vasculature of a Tek-Cre::Ai47 transgenic mouse imaged by high-definition fMOST (HD-fMOST) [31]. The transgenic mouse is a hybrid of a Tek-Cre transgenic mouse and an Ai47 transgenic mouse. In this mouse, the epithelial cells in the vascular wall were labelled with green-fluorescent protein (GFP). Then the mouse brain was removed and imaged continuously using the HD-fMOST at a resolution of $0.32\mu m \times 0.32\mu m \times 1\mu m$ voxels. The size of the uncompressed image dataset exceeded 20 terabytes (16-bit pixel depth) with more than 12,000 coronal sections.

We resampled the resolution of the two datasets to 1 μm / pixel and randomly cropped 82 volumes of size $200 \times 200 \times 200$ pixels from Dataset 1 and 100 volumes of size $160 \times 160 \times 160$ pixels from Dataset 2. We used the segmentation editor module of Amira software to annotate the vessels in the volumes manually. The original volumes and annotated volumes were used to train and test our proposed network.

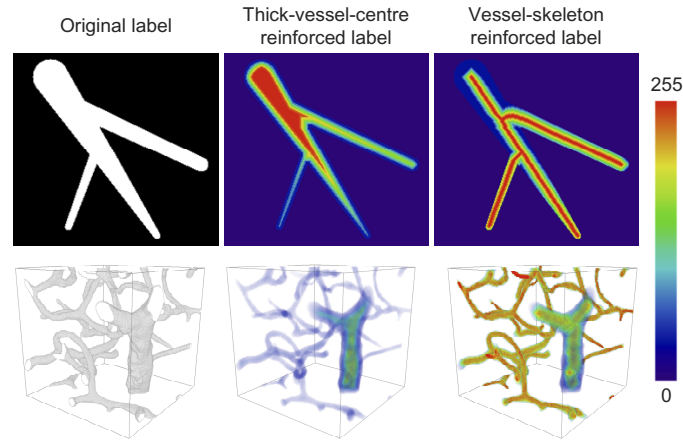## 2.2. Multi-perspective label generation

Based on the annotated vessel binary labels, two different perspectives of labelling data are generated using morphological operation. Namely, the thick-vessel-centre reinforced label and the vessel skeleton reinforced label (see Fig. 2). Thick-vessel-center reinforced labels are generated by performing continuous erosion operations on the original vessel binary labels until all foreground pixels are eroded, or the specified number of erosion rounds is reached. During the erosion process, the pixels eroded in each round are recorded in a blank image according to pixel coordinates and given a different grey value. The more rounds of erosion, the higher the grey value of the corresponding eroded pixels. Since the erosion operation will erode from the edges of the vessels inward, more erosion rounds are needed to erode the thick vessels than the thin vessels. The pixel values in the centre regions of thick vessels are higher than those in the other regions. The generation process of the vessel skeleton reinforced labels firstly extracts the vascular centerlines using 3D skeletonization and then performs successive dilation operations on the centerlines. For each round of dilation, the dilated pixels are given a different grey value, and the more rounds of dilation, the lower value of the corresponding pixels. During dilation, the original vessel binary label is used as a mask, and dilation is stopped when all pixels in the mask are covered by the dilated pixels. Due to the dilation operation from the centerlines of vessels outward, the pixel values of the centre region of all vessels are brightest in generated labels, regardless of the thickness of vessels.

The two different perspective labels use the grey-scale difference to mark the centre of thick vessels as well as the skeletal regions of vessels. In the training stage, these labels are used to train the network, allowing the network to focus on these hard-to-segment regions.
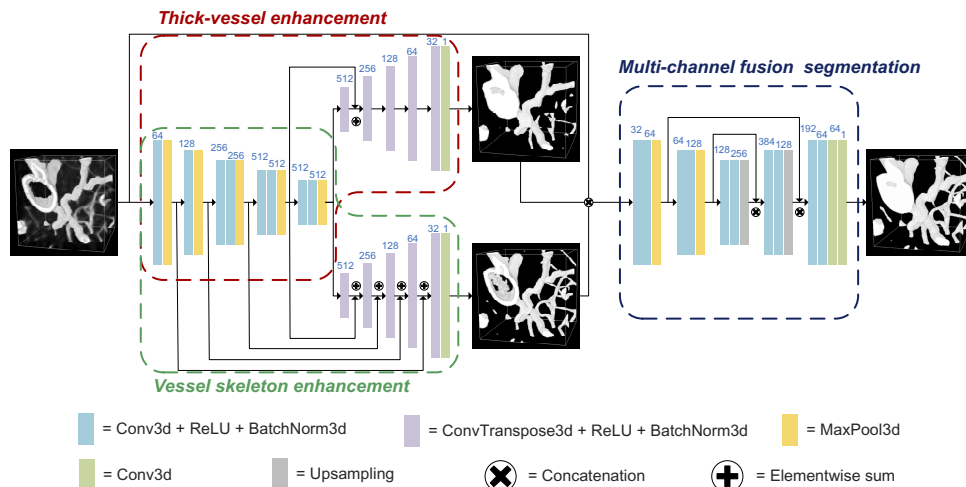
2.3. Network architecture

We designed a three-stage network, as shown in Fig. 3. The architecture contains three sub-networks: thick-vessel enhancement network, vessel skeleton enhancement network, and multi-channel fusion segmentation network.

Fig. 2. 2D/3D exemplar results of multi-perspective labels generation process. Gray images are transformed into pseudo-colour images for better visualization.

Fig. 3. The architecture of the proposed three-stage framework. The framework contains three sub-networks: thick-vessel enhancement network, vessel skeleton enhancement network, and multi-channel fusion segmentation network.

All three sub-networks use an encoder-decoder U-shape network structure. The vessel features are extracted from the original image by the encoder network, and as the network gets deeper, more high-level image features are extracted. The decoder network fuses and decodes the features at different scales and finally outputs a prediction map at the original resolution. Since the function of each sub-network is different, the structure of the sub-networks differs in some details.

Both the thick-vessel enhancement network and the vessel skeleton enhancement network use the 3D fully convolutional networks (FCNs) [32]. This network structure extracts features at different scales by sampling five times. The deeper the sampling, the larger the receptive field is, and more global semantic information is extracted. Then the probability map of each pixel in the image is output by five times of upsampling. Since the role of the thick-vessel enhancement network is to pre-segment the thick vessels, especially to identify and fill the cavities in the centre of thick vessels, these features belong to the global semantic information. Therefore, the thick-vessel enhancement network reads only the last two scales of the feature layers for decoding and predicting the probability map. The vessel skeleton enhancement network needs to identify all vascular skeletons. In some thin vessels, the skeleton features are not obvious due to the low intensity of fluorescence signals. That means the network needs not only global semantic information but also shallow (pixel) information. Therefore, in the decoding stage, we utilize all feature layers for decoding and predicting the probability map.

The multi-channel fusion segmentation network uses a modified 3D U-net [33]. This network reads the outputs of the two previous sub-networks and the original vessel volumes to predict the final segmentation results. As most of the vessels have been pre-segmented by the first two sub-networks, we only perform three downsampling to extract features to reduce the network size.

2.4. Training strategy and loss function

The three sub-networks are trained separately. The thick-vessel enhancement network reads the original vessel volumes and is trained using thick-vessel-centre reinforced labels. In the training stage, only pixels with grey-scale values above a given threshold (pixels in the centre of thick vessels) and background pixels are used for the loss calculation and backpropagation. Similarly, the vessel skeleton enhancement network reads the original volumes and the vessel skeleton reinforced labels for training. Only the pixels in the vascular skeleton region and the background pixels participate in the loss calculation and back propagation during the training process. The multi-channel fusion segmentation network reads the original images, as well as the predicted probability maps generated by the first two sub-networks, and is trained using the original binary labels to predict the final segmentation results.

The soft-Dice loss is adopted in the training stage:

$$loss = 1 - \sum_{i=0}^{W} \sum_{j=0}^{H} \sum_{k=0}^{D} m_{ijk} \left(\frac{2 \times y_{ijk} \times \bar{y}_{ijk}}{y_{ijk}^2 + \bar{y}_{ijk}^2} \right), \tag{1}$$

where y_{ijk} is the truth value of the pixel at (i, j, k) in the 3D volume. 1 denotes that the pixel belongs to the vessels and 0 denotes that the pixel belongs to the background. $\bar{y}_{ijk}$ is the predicted value. We use a sigmoid function to constrain $\bar{y}_{ijk}$ between 0 and 1. m_{ijk} is a mask that determines which pixels in the labels participate in the loss calculation. Since only some pixels in the labels participate in the loss calculation and back propagation when the first two networks are trained, a mask is needed to filter these pixels that are involved in the loss calculation. The size of m_{ijk} is the same as the label size. The value of m_{ijk} is 1 when the pixel at (i, j, k) are involved in the loss calculation, otherwise, the value of m_{ijk} is 0. In the third sub-network, since all pixels are involved in the calculation of the loss for the prediction of the final vessel segmentation results, the value of m_{ijk} is 1.

2.5.   Implementation details

We use PyTorch to implement our framework. All experiments were performed on a computer with Intel Core CPU i9-9900k, 64GB RAM, and NVIDIA RTX 3090 GPU with 24GB video memory. In the training stage, The gradient descent algorithm is set as 'Adam'. The learning rate is 0.0001, and the batch size is 2. The dataset was divided into training, validation, and testing sets. The proportion of the test set to the whole datasets are 12% (Dataset 1) and 20% (Dataset 2), respectively. In each epoch of the training process, volumes with a fixed size of $128 \times 128 \times 128$ pixels were fed into the network. These volumes were randomly cropped from the original volumes and randomly changed the order of the x, y, z dimensions (random transpose) to augment the size of the training dataset. The training and validation loss curves are shown in Figs. S1–2.

2.6.   Evaluation metrics

The final output of our network is a probability map. A binary segmentation result can be generated by the threshold value of 0.5. Pixels with a value greater than or equal to 0.5 are the foreground, and the other pixels are the background. We used four overlap-based metrics: *Accuracy*, *Precision*, *Recall*, and F_1-*score* (*Dice*). We also used two morphological similarity-based metric: Hausdorff Distance (*Hausdorff*), as in Tahir et al. [19] and centerlineDice (*clDice*) [9]. The evaluation metrics are described below:

Let *TP* notes the number of vessel pixels correctly detected, *FP* notes the number of vessel pixels incorrectly detected, *TN* notes the number of non-vessel pixels correctly detected, *FN* denotes the number of non-vessel pixels incorrectly detected. Then, *Accuracy* is defined as *Acc* = (*TP*+*TN*)/(*TP*+*TN*+*FP*+*FN*). *Precision* is defined as *Prec* = *TP*/(*TP*+*FN*). *Recall* is defined as *Rec* = *TN*/(*TN*+*FP*). F_1-*score* or *Dice* is the harmonic mean of *Precision* and *Recall*, and it is defined as F_1 = (2*TP*)/(2*TP*+*FP*+*FN*).

Hausdorff measures how far two subsets of a metric space are from each other. the Hausdorff distance between A and B is defined as $H(A, B) = \max(h(A, B), h(B, A))$, where $h(A, B)$ is the Hausdorff distance from set A to set B, defined as $h(A, B) = \max_{a \in A}(\min_{b \in B}(d(a, b)))$, where a and b are points of sets A and B respectively, and $d(a, b)$ is any metric (e.g. Euclidian distance) between a and b. *clDice* measures the topological similarity of the segmentation results and the ground truths. The specific introduction of *clDice* can be found in [9]. The lower the score of *Hausdorff* and the higher the score of *clDice*, the more the similarity between the topology of the segmentation results and the topology of the ground truths.

3.   Results

3.1.   Visualization results

We visualize the pre-segmentation results of the first two sub-networks and the final segmentation results, and compare them with the ground truths. The result are shown Fig. 4. As mentioned above, the thick-vessel enhancement network focuses on predicting the central region of the thick vessels in the volumes, and the cavities in thick vessels can be filled effectively, but the high-level features extraction ability of the network structure makes it ineffective in predicting fine vessels or low fluorescent signal regions. On the other hand, the vessel skeleton enhancement network deals with the segmentation of the skeletal region of all vessels in the image. The adopted network structure can effectively capture the details in the volume. The network can accurately segment the vessels at some weak fluorescence signal regions, thus preserving the vascular topology connectivity. However, it does not work well for cavities in thick vessels. The multi-channel fusion segmentation network combines the prediction results of the first two sub-network with the original volumes to output the final vessel segmentation results. The results effectively predict the cavities in thick vessels and the skeletal region of vessels. It can be seen

that the final segmentation results are very close to the gold truths. The more visualization results are shown in Figs. S3–4.
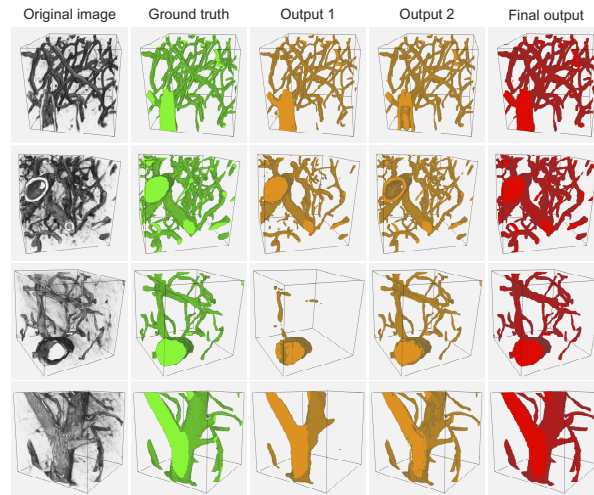

Fig. 4. Four examples of vessel segmentation results. Output 1 and Output 2 are the pre-segmentation results of the thick-vessel enhancement network and vessel skeleton enhancement network. The first and second rows are Dataset 1 (volume size: $192 \times 192 \times 192$ μm), and the third to fourth rows are Dataset 2 (volume size: $160 \times 160 \times 160$ μm).

We also compare the proposed method with other similar methods, namely OTSU [34], VesSAP [22], 3D-Unet [33] and Vnet [35]. OTSU is a traditional segmentation method, and the other three are deep learning-based methods. The results are shown in Fig. 5. It can be seen that OTSU and VesSAP are difficult to deal with the cavities in the centre of thick vessels as well as the weak fluorescence signal vessels, and the predicted results are largely different from the ground truths. Although 3D-Unet and Vnet are able to fill some of the cavities in the thick vessels, they still cannot segment well at some of the thin vessels, resulting in a discontinuous vascular topology. In addition, in Dataset 2, 3D-Unet and Vnet produced large false segmentation around the thick vessels. They predicted some background pixels as thick vessels, resulting in inaccurate segmentation. Only our method can effectively segment both the cavities in thick vessels and the weak fluorescence signal vessels. The segmentation effect is significantly better than the other four methods. We also selected some detailed regions from the test set and showed the segmentation results of our method and 3D-Unet (see Fig. 6). It can be seen that our method has an obvious advantage both in terms of the precision of pixel segmentation and vascular topology preservation.

3.2. Quantitative results

In this subsection, we quantitatively evaluate the effectiveness of our method and compare the results with the other four methods. The evaluation metrics were introduced in the previous section. Table 1 shows the final results for each approach on different metrics. The best scores have been bolded. As can be seen, the quantitative results are consistent with the visualization results, and our method achieves almost the best results. Our approach employs two sub-networks that target pre-segmentation of the central region of the thick vessels and the skeletal region of the vessels. The network can fill the cavities in the centre of thick vessels and identify and segment the weak signals in the fine vessels, resulting in a high *recall*. On the other hand, as the background pixels have participated in the loss calculation during the training of all three sub-networks, the
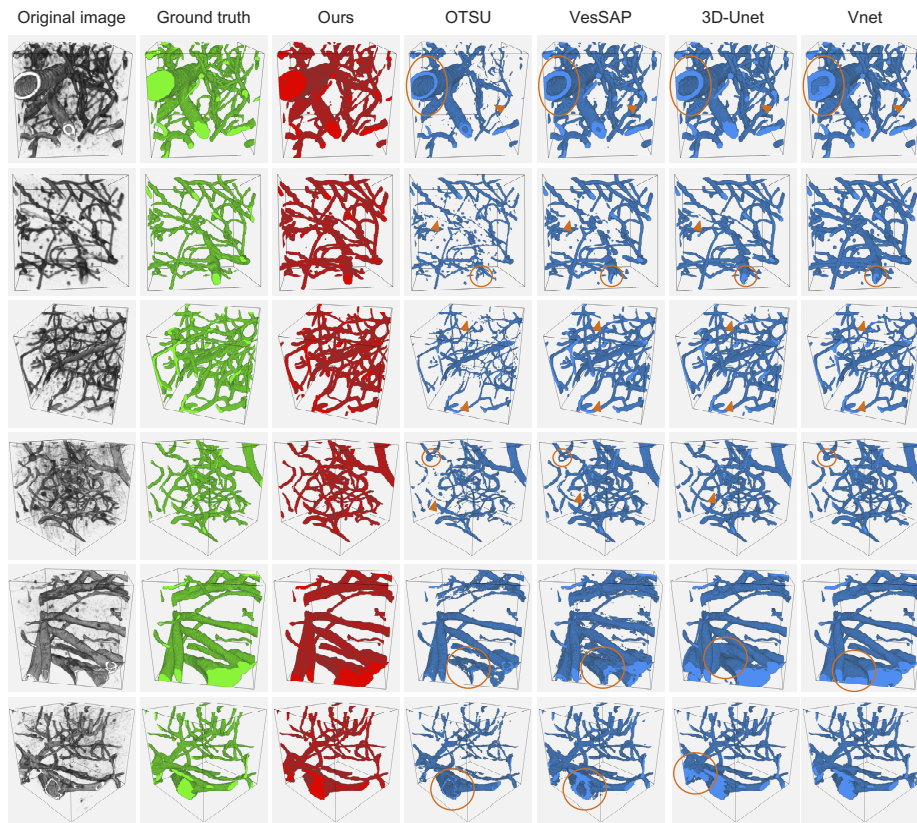
Fig. 5. Comparative segmentation results using different methods on six examples. The first to third rows are Dataset 1 (volume size: $192 \times 192 \times 192 \ \mu$m), and the fourth to sixth rows are Dataset 2 (volume size: $160 \times 160 \times 160 \ \mu$m). Some false segmentations are highlighted with orange arrows and circles.
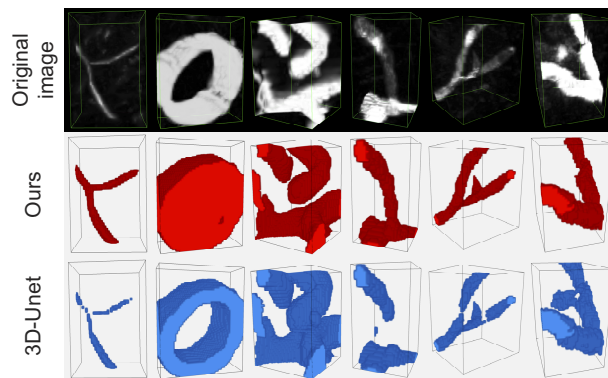
Fig. 6. Six enlarged regions from the test set and the corresponding segmentation results generated by our proposed method and 3D-Unet. From left to right, the volume sizes are 37×56×29 μm, 50×64×21 μm, 38×55×38 μm, 24×43×24 μm, 37×44×33 μm, 30×48×40 μm, respectively.

segmentation results also have better suppression of the noise in the background, resulting in a high *precision*.

Table 1. Quantitative comparison results on different datasets. The detailed numerical data can be found in Data File 1 [36] and Data File 2 [37].

Method	Dataset 1						Dataset 2					
	Acc	*Prec*	*Recall*	*F_1-score*	*Hausdorff*	*clDice*	*Acc*	*Prec*	*Recall*	*F_1-score*	*Hausdorff*	*clDice*
OTSU	96.15	**99.11**	42.81	58.94	31.63	84.72	96.89	**99.22**	33.73	49.47	37.26	48.65
VesSAP	98.06	87.86	84.63	85.17	27.60	96.06	98.68	90.90	86.57	88.11	40.24	91.64
3D-Unet	97.91	86.05	87.30	85.40	26.28	96.99	99.07	91.90	95.01	93.08	30.11	94.67
Vnet	98.03	89.89	81.33	84.25	28.99	94.58	99.43	95.59	93.25	94.26	36.44	94.99
Ours	**98.69**	87.03	**92.44**	**89.40**	**24.93**	**97.88**	99.74	97.26	**96.73**	**96.96**	**29.98**	**98.66**

To further demonstrate the effect of our method, we performed a quantitative analysis of the pixels in the thick-vessel region and the micro-vessel region separately. These regions make segmentation difficult and error-prone. Based on the ground truths, we extracted the centerlines of vessels and calculated the vessel diameters. Then we extracted the vessels with diameters greater than 10 pixels and less than 3 pixels, representing the thick vessel regions and the micro-vessel regions, respectively. For each extracted region, a 3-pixel range is assigned to cover non-vessel pixels to calculate the metrics (*TN* and *FN*) (see Fig. 7). We calculated *Acc*, *Prec*, *Recall*, and *F_1-score* separately and compared the results with the other four methods. Table 2 show the comparison results. As can be seen, our approach also achieved the best results.
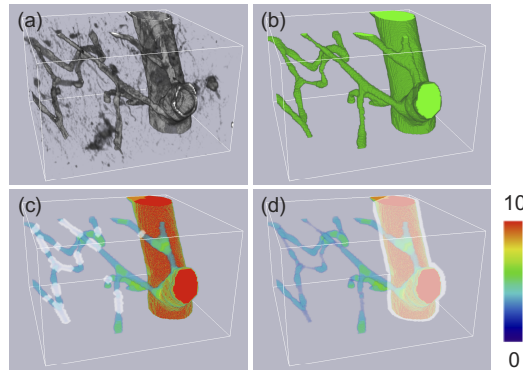
Fig. 7. Vessel pixels for quality evaluation on thick-vessel segmentation and micro-vessel segmentation. (a) The original volume. (b) The ground truth annotation. (c) The defined micro-vessel pixels (white regions). (d) The defined thick-vessel pixels thick-vessel. Volume size: $147 \times 88 \times 141$ μm.

3.3. Ablation studies

Our method proposed a multi-perspective label based vessel segmentation framework and designed a three-stage network structure for training and predicting. In this subsection, we designed two ablation experiments to demonstrate that the proposed method effectively improves vessel segmentation. Both ablation experiments were performed on Dataset 1.

First, we verified the effect of each sub-network alone for vessel segmentation and compared it with the three-stage network. As all three sub-networks are end-to-end segmentation networks, we trained the networks by using the original volumes and annotated vessel binary labels. After training, the networks were validated on the test set. We also use 'OR' operation instead of

Table 2. Quantitative comparison results on micro-vessel and thick-vessel segmentation.

Type	Method	Dataset 1				Dataset 2			
		Acc	*Prec*	*Recall*	*F_1-score*	*Acc*	*Prec*	*Recall*	*F_1-score*
Micro-vessel	OTSU	68.55	**99.33**	27.21	41.18	79.28	**99.88**	41.32	57.85
	DVN	83.28	80.33	84.58	81.17	94.91	92.92	93.03	92.75
	3D-Unet	81.86	78.85	85.33	80.16	95.51	96.56	90.78	93.39
	Vnet	80.10	83.94	70.96	74.31	94.48	95.33	89.23	91.85
	Ours	**83.96**	78.45	**88.52**	**82.67**	**96.52**	95.28	**94.98**	**95.01**
Thick-vessel	OTSU	65.90	**99.87**	45.61	60.56	55.49	**99.97**	27.05	41.40
	DVN	81.97	94.02	76.72	83.02	85.59	93.76	83.30	87.45
	3D-Unet	87.52	91.97	89.10	89.77	97.42	98.64	97.17	97.86
	Vnet	87.94	93.39	87.56	90.06	95.88	98.30	95.24	96.59
	Ours	**92.73**	93.89	**94.29**	**93.98**	**97.68**	98.38	**97.86**	**98.09**

sub-network 3 to combine the pre-segmentation results of the first two sub-networks. The results are shown in Table 3. It can be seen that when each sub-network is used individually for vessel segmentation, the segmentation effects of the sub-networks are not as good as that of the three-stage network. The results indicate that three sub-networks all play a role in the three-stage framework. The three sub-networks pre-segment the cavities in thick vessels, the skeletal regions of vessels, and fuse the pre-segmentation results, respectively. The three cooperate with each other to improve the effect of vessel segmentation.

Table 3. Ablation analysis of sub-networks of the proposed framework on Dataset 1. The detailed numerical data can be found in Data File 3 [38].

Module	*Acc*	*Prec*	*Recall*	*F_1-score*	*Hausdorff*	*clDice*
Sub-network 1	98.00	89.55	80.52	83.76	28.03	94.77
Sub-network 2	97.77	83.33	90.46	85.76	27.60	92.49
Sub-network 3	98.19	90.22	82.92	85.30	27.18	96.49
Sub-network 1 ∪ 2	97.34	**97.65**	60.88	74.47	25.47	96.45
Three-stage network	**98.69**	87.03	**92.44**	**89.40**	**24.93**	**97.88**

Second, our proposed multi-perspective labels combined with the three-stage segmentation framework can also enhance the existing segmentation network. We replace the multi-channel fusion segmentation network with VesSAP, an end-to-end fully convolutional neural network. We set the number of input channels of VesSAP to three and feed the pre-segmentation results of the first two sub-networks and the original volumes into VesSAP for training and predicting. The results are compared with those of VesSAP segmented independently (see Table 4). When the pre-segmentation results were used, the segmentation effect of VesSAP was significantly improved. This indicates that our proposed multi-perspective labels enable the network to focus on the regions highlighted by the labels. The pre-segmented results could extract the hard-to-segment features in the original images, so that the effect of vessel segmentation can be improved.
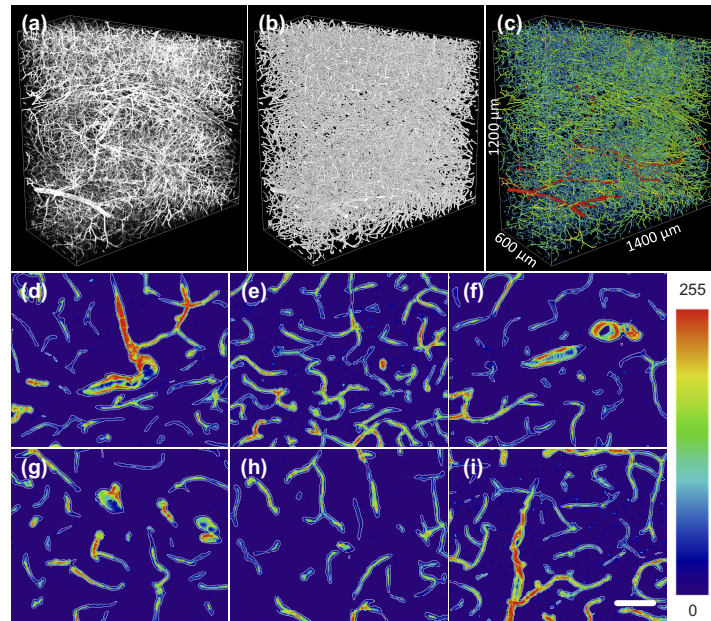
3.4. Large-scale vessel segmentation

We extend our method to segment large-scale vascular data. We cropped a volume of size 1,400 × 1,200 × 600 pixels from Dataset 1. Based on our previous work of TDat [40], the volume is split into 1,008 overlapping 3D blocks, each with a fixed size of 128 × 128 × 128 pixels. The 3D

Table 4. Ablation study of multi-perspective labels (MPLs) on Dataset 1. The detailed numerical data can be found in Data File 4 [39].

Module	*Acc*	*Prec*	*Recall*	*F_1-score*	*Hausdorff*	*clDice*
VesSAP	98.06	**87.86**	84.63	85.17	27.60	96.06
VesSAP + MPLs	**98.35**	82.76	**93.69**	**87.49**	**25.41**	**97.81**

blocks are segmented and merged. Pixels in the overlapping area between adjacent small blocks are fused by the 'OR' operation. After segmentation, we used some morphological operations, including 2D and 3D hole filling, removing small connected domains, in order to reconstruct the vessel network more accurately. The total segmentation and reconstruction time is 11 minutes. The result is shown in Fig. 8 and Visualization 1. We show the original vascular volume, the segmented volume, and the reconstructed vascular network based on the segmented volume. In addition, we also randomly selected six regions to show the segmentation results in 2D projection. It can be seen that our method can perform fast and accurate segmentation for a large scale of vascular image data.


Fig. 8. Vessel segmentation on large-scale images. (a) The original volume. (b) The corresponding segmentation result. (c) The reconstructed vascular network based on the segmented volume. (d)-(i) 2D max-intensity projection (thickness: 20 μm) of some regions randomly selected from the original volume. The white curves represent the contour of the segmentation volume. The original grey images are transformed into pseudo-colour images for better visualization. Scale bar, 50 μm.

4. Discussion

We propose a three-stage network framework based on multi-perspective labels. It is used to segment whole-brain fluorescence vascular images. Unlike other vascular imaging methods, fluorescently labelled vessels cause cavities in the centre of thick vessels. In addition, the weak fluorescence signal in fine vessels resulted in hard-to-segment, so that the topological connectivity

of vessels cannot be maintained. Our proposed method segments these 'hard' regions by two sub-networks. Compared with other methods, our method achieved state-of-the-art results.

The reason that causes the deep learning network to be challenging to segment effectively in thick vessels and fine vessel skeleton regions is that the vessel diameter distribution is extremely uneven. Most of vessels are capillaries. The percentage of the large and very thin vessel is very small, so these hard-to-segment features account for a very small percentage of the training samples. During the training process, the loss in these regions is overwhelmed by the loss in most of the easy-to-segment regions. The network cannot effectively capture the features of these hard-to-segment regions, resulting in poor segmentation. Our proposed multi-perspective labels can screen these hard-to-segment regions and only calculate the loss of these regions during training, thus allowing the network to focus on learning the features of these regions. In fact, if the training dataset is purposefully selected or if the training dataset contains a high percentage of hard-to-segment samples, the effect of the networks on segmenting these regions will also improve. It can be noted that Vnet and 3D-Unet also work well on Dataset 2 for the segmentation of cavities in thick vessels. It is because the transgenic fluorescent labelling method only labels the thin epithelial cells in vessels, leaving most of vessels in images containing cavities (see Fig. 9). This results in a high percentage of 'cavity' features in the training samples, so the networks can easily learn. In contrast, in the fluorescent dye-labelled vessel images (Dataset 1), the cavities in the fine vessels are not obvious, and only the thick vessels contain cavities, so the network has difficulty learning 'cavity' features.
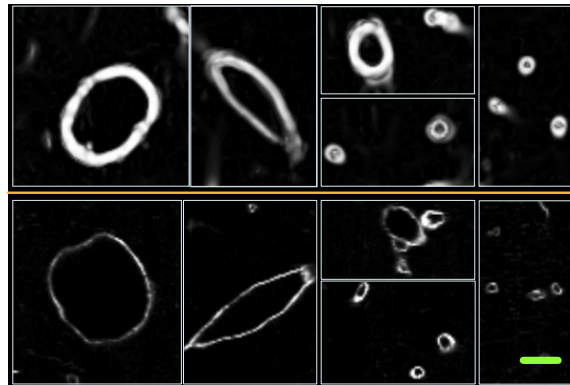
Fig. 9. Cross-section of blood vessels with different vessel diameters from two datasets. From top to bottom: fluorescent dye-labelled vessels (Dataset 1) and the transgenic-labelled vessels (Dataset 2). Scale bar, 20 μm.

Although our method achieves good results on segmentation effects, our approach has some weaknesses. The biggest weakness is the complex network structure. In theory, the entire framework could be trained end-to-end. However, the biggest problem is that end-to-end training requires three sub-networks to be read into the GPU memory at the same time for training, which consumes GPU memory resources very much. To solve this problem, we trained the three sub-networks separately, consuming more time and resources than the one-stage end-to-end segmentation network. In prediction, a volume needs to be read three times to complete the final segmentation. However, since the features of data acquired by the same imaging modality are very similar, once the network is trained on a dataset, the trained network can be used directly to segment other data acquired by the same imaging modality, thus avoiding retraining of the network. In addition, our method takes only 10 minutes to segment 1 GB of volume, which is an acceptable sacrifice in time compared to the improvement in segmentation effect.

5. Conclusion

This paper presents a multi-perspective label-based three-stage deep learning framework to segment cerebral vasculature in whole-brain fluorescence images. The proposed framework uses morphological operations to screen the thick vessel centre and vessel skeleton regions in binary annotated labels to generate two different labels. These two labels mark the regions of vessels that are difficult to segment. Then we use two sub-networks to pre-segment the vessels based on these two labels respectively, and the third sub-networks to fuse the pre-segmented results to precisely segment the vessels. Experimental results show that our method achieves state-of-the-art performance. In future works, we will extend our approach to whole-brain datasets, enabling segmentation and reconstruction of all vascular networks in the whole brain.

Disclosures. The authors declare no conflicts of interest.

Data availability. Data underlying the results presented in this paper are available in Data File 1 [36], Data File 2 [37], Data File 3 [38], Data File 4 [39].

Supplemental document. See Supplement 1 for supporting content.

References

1. V. Muoio, P. Persson, and M. Sendeski, "The neurovascular unit–concept review," Acta Physiol. **210**(4), 790–798 (2014).
2. B. J. Andreone, B. Lacoste, and C. Gu, "Neuronal and vascular interactions," Annu. Rev. Neurosci. **38**(1), 25–46 (2015).
3. K. Kisler, A. R. Nelson, A. Montagne, and B. V. Zlokovic, "Cerebral blood flow regulation and neurovascular dysfunction in alzheimer disease," Nat. Rev. Neurosci. **18**(7), 419–434 (2017).
4. B. V. Zlokovic, "Neurovascular pathways to neurodegeneration in alzheimer's disease and other disorders," Nat. Rev. Neurosci. **12**(12), 723–738 (2011).
5. M. D. Sweeney, K. Kisler, A. Montagne, A. W. Toga, and B. V. Zlokovic, "The role of brain vasculature in neurodegenerative disorders," Nat. Neurosci. **21**(10), 1318–1331 (2018).
6. J. Wu, Y. He, Z. Yang, C. Guo, Q. Luo, W. Zhou, S. Chen, A. Li, B. Xiong, T. Jiang, and H. Gong, "3d braincv: simultaneous visualization and analysis of cells and capillaries in a whole mouse brain with one-micron voxel resolution," NeuroImage **87**, 199–208 (2014).
7. C. Kirst, S. Skriabine, A. Vieites-Prado, T. Topilko, P. Bertin, G. Gerschenfeld, F. Verny, P. Topilko, N. Michalski, M. Tessier-Lavigne, and N. Renier, "Mapping the fine-scale organization and plasticity of the brain vasculature," Cell **180**(4), 780–795.e25 (2020).
8. X. Ji, T. Ferreira, B. Friedman, R. Liu, H. Liechty, E. Bas, J. Chandrashekar, and D. Kleinfeld, "Brain microvasculature has a common topology with local differences in geometry that match metabolic load," Neuron **109**(7), 1168–1187.e13 (2021).
9. S. Shit, J. C. Paetzold, A. Sekuboyina, I. Ezhov, A. Unger, A. Zhylka, J. P. Pluim, U. Bauer, and B. H. Menze, "cldice-a novel topology-preserving loss function for tubular structure segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, (2021), pp. 16560–16569.
10. S. Moccia, E. De Momi, S. El Hadji, and L. S. Mattos, "Blood vessel segmentation algorithms–review of methods, datasets and evaluation metrics," Comput. Methods Programs Biomed. **158**, 71–91 (2018).
11. D. Lesage, E. D. Angelini, I. Bloch, and G. Funka-Lea, "A review of 3d vessel lumen segmentation techniques: Models, features and extraction schemes," Med. Image Anal. **13**(6), 819–845 (2009).
12. D. Jia and X. Zhuang, "Learning-based algorithms for vessel tracking: A review," Comput. Med. Imaging Graph. **89**, 101840 (2021).
13. A. F. Frangi, W. J. Niessen, K. L. Vincken, and M. A. Viergever, "Multiscale vessel enhancement filtering," in *International conference on medical image computing and computer-assisted intervention*, (Springer, 1998), pp. 130–137.
14. Y. Zhao, Y. Zheng, Y. Liu, Y. Zhao, L. Luo, S. Yang, T. Na, Y. Wang, and J. Liu, "Automatic 2-d/3-d vessel enhancement in multiple modality images using a weighted symmetry filter," IEEE Trans. Med. Imaging **37**(2), 438–450 (2018).
15. Y. Zeng, Y. Zhao, P. Tang, M. Liao, Y. Liang, S. Liao, and B. Zou, "Liver vessel segmentation and identification based on oriented flux symmetry and graph cuts," Comput. Methods Programs Biomed. **150**, 31–39 (2017).
16. Y. Shang, R. Deklerck, E. Nyssen, A. Markova, J. De Mey, X. Yang, and K. Sun, "Vascular active contour for vessel tree segmentation," IEEE Trans. Biomed. Eng. **58**(4), 1023–1032 (2011).

17. Y. Cheng, X. Hu, J. Wang, Y. Wang, and S. Tamura, "Accurate vessel segmentation with constrained b-snake," IEEE Trans. on Image Process. **24**(8), 2440–2455 (2015).

18. A. Goyal, J. Lee, P. Lamata, J. van den Wijngaard, P. van Horssen, J. Spaan, M. Siebes, V. Grau, and N. P. Smith, "Model-based vasculature extraction from optical fluorescence cryomicrotome images," IEEE Trans. Med. Imaging **32**(1), 56–72 (2013).

19. W. Tahir, S. Kura, J. Zhu, X. Cheng, R. Damseh, F. Tadesse, A. Seibel, B. S. Lee, F. Lesage, S. Sakadžic, D. A. Boas, and L. Tian, "Anatomical modeling of brain vasculature in two-photon microscopy by generalizable deep learning," BME Front. **2021**, 1–12 (2021).

20. M. Haft-Javaherian, L. Fang, V. Muse, C. B. Schaffer, N. Nishimura, and M. R. Sabuncu, "Deep convolutional neural networks for segmenting 3d in vivo multiphoton images of vasculature in alzheimer disease mouse models," PLoS One **14**(3), e0213539 (2019).

21. R. Damseh, P. Pouliot, L. Gagnon, S. Sakadzic, D. Boas, F. Cheriet, and F. Lesage, "Automatic graph-based modeling of brain microvessels captured with two-photon microscopy," IEEE J. Biomed. Health Inform. **23**(6), 2551–2562 (2019).

22. M. I. Todorov, J. C. Paetzold, O. Schoppe, G. Tetteh, S. Shit, V. Efremov, K. Todorov-Völgyi, M. Düring, M. Dichgans, M. Piraud, B. Menze, and A. Ertürk, "Machine learning analysis of whole mouse brain vasculature," Nat. Methods **17**(4), 442–449 (2020).

23. G. Tetteh, V. Efremov, N. D. Forkert, M. Schneider, J. Kirschke, B. Weber, C. Zimmer, M. Piraud, and B. H. Menze, "Deepvesselnet: Vessel segmentation, centerline prediction, and bifurcation detection in 3-d angiographic volumes," Front. Neurosci. **14**, 1285 (2020).

24. Q. Yan, B. Wang, W. Zhang, C. Luo, W. Xu, Z. Xu, Y. Zhang, Q. Shi, L. Zhang, and Z. You, "Attention-guided deep neural network with multi-scale feature fusion for liver vessel segmentation," IEEE J. Biomed. Health Inform. **25**(7), 2629–2642 (2021).

25. K. Li, X. Qi, Y. Luo, Z. Yao, X. Zhou, and M. Sun, "Accurate retinal vessel segmentation in color fundus images via fully attention-based networks," IEEE J. Biomed. Health Inform. **25**(6), 2071–2081 (2021).

26. D. Wang, A. Haytham, J. Pottenburgh, O. Saeedi, and Y. Tao, "Hard attention net for automatic retinal vessel segmentation," IEEE J. Biomed. Health Inform. **24**(12), 3384–3396 (2020).

27. Z. Wang, X. Jiang, J. Liu, K.-T. Cheng, and X. Yang, "Multi-task siamese network for retinal artery/vein separation via deep convolution along vessel," IEEE Trans. Med. Imaging **39**(9), 2904–2919 (2020).

28. Z. Yan, X. Yang, and K.-T. Cheng, "A three-stage deep learning model for accurate retinal vessel segmentation," IEEE J. Biomed. Health Inform. **23**(4), 1427–1436 (2019).

29. Z. Yan, X. Yang, and K.-T. Cheng, "A skeletal similarity metric for quality evaluation of retinal vessel segmentation," IEEE Trans. Med. Imaging **37**(4), 1045–1057 (2018).

30. H. Gong, D. Xu, J. Yuan, X. Li, C. Guo, J. Peng, Y. Li, L. A. Schwarz, A. Li, B. Hu, B. Xiong, Q. Sun, Y. Zhang, J. Liu, Q. Zhong, T. Xu, S. Zeng, and Q. Luo, "High-throughput dual-colour precision imaging for brain-wide connectome with cytoarchitectonic landmarks at the cellular level," Nat. Commun. **7**(1), 12142 (2016).

31. Q. Zhong, A. Li, R. Jin, D. Zhang, X. Li, X. Jia, Z. Ding, P. Luo, C. Zhou, C. Jiang, Z. Feng, Z. Zhang, H. Gong, J. Yuan, and Q. Luo, "High-definition imaging using line-illumination modulation microscopy," Nat. Methods **18**(3), 309–315 (2021).

32. J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, (2015), pp. 3431–3440.

33. Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3d u-net: learning dense volumetric segmentation from sparse annotation," in *International conference on medical image computing and computer-assisted intervention*, (Springer, 2016), pp. 424–432.

34. N. Otsu, "A threshold selection method from gray-level histograms," IEEE Trans. Syst., Man, Cybern. **9**(1), 62–66 (1979).

35. F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *2016 fourth international conference on 3D vision (3DV)*, (IEEE, 2016), pp. 565–571.

36. Y. Li, T. Ren, J. Li, X. Li, and A. Li, "Data file 1," figshare (2022). Retrieved: https://opticapublishing.figshare.com/s/e76d6a7d8ef551a9c7ca.

37. Y. Li, T. Ren, J. Li, X. Li, and A. Li, "Data file 2," figshare (2022). Retrieved: https://opticapublishing.figshare.com/s/4e27cb1d3751639deb6b.

38. Y. Li, T. Ren, J. Li, X. Li, and A. Li, "Data file 3," figshare (2022). Retrieved: https://opticapublishing.figshare.com/s/80516ead2167486ce706.

39. Y. Li, T. Ren, J. Li, X. Li, and A. Li, "Data file 4," figshare (2022). Retrieved: https://opticapublishing.figshare.com/s/26c26a954b812a466104.

40. Y. Li, H. Gong, X. Yang, J. Yuan, T. Jiang, X. Li, Q. Sun, D. Zhu, Z. Wang, Q. Luo, and A. Li, "Tdat: an efficient platform for processing petabyte-scale whole-brain volumetric images," Front. Neural Circuits **11**, 51 (2017).